

# Testing Formula Satisfaction\*

Eldar Fischer<sup>†</sup>Yonatan Goldhirsh<sup>‡</sup>Oded Lachish<sup>§</sup>

## Abstract

We study the query complexity of testing for properties defined by read once formulae, as instances of *massively parametrized properties*, and prove several testability and non-testability results. First we prove the testability of any property accepted by a Boolean read-once formula involving any bounded arity gates, with a number of queries exponential in  $\epsilon$  and independent of all other parameters. When the gates are limited to being monotone, we prove that there is an *estimation* algorithm, that outputs an approximation of the distance of the input from satisfying the property. For formulae only involving And/Or gates, we provide a more efficient test whose query complexity is only quasipolynomial in  $\epsilon$ . On the other hand we show that such testability results do not hold in general for formulae over non-Boolean alphabets; specifically we construct a property defined by a read-once arity 2 (non-Boolean) formula over alphabets of size 4, such that any  $1/4$ -test for it requires a number of queries depending on the formula size.

## 1 Introduction

*Property Testing* deals with randomized approximation algorithms that operate under low information situations. The definition of a property testing algorithm uses the following components: A set of *objects*, usually the set of strings  $\Sigma^*$  over some alphabet  $\Sigma$ ; a notion of a single *query* to the input object  $w = (w_1, \dots, w_n) \in \Sigma^*$ , which in our case would consist of either retrieving the length  $|w|$  or the  $i$ 'th letter  $w_i$  for any  $i$  specified by the algorithm; and finally a notion of *farness*, a normalized distance, which in our case will be the Hamming distance —  $\mathbf{farness}(w, v)$  is defined to be  $\infty$  if  $|w| \neq |v|$  and otherwise it is  $|\{i : w_i \neq v_i\}|/|v|$ .

Given a *property*  $P$ , that is a set of objects  $P \subseteq \Sigma^*$ , an integer  $q$ , and a farness parameter  $\epsilon > 0$ , an  $\epsilon$ -test for  $P$  with *query complexity*  $q$  is an algorithm that is allowed access to an input object only through queries, and distinguishes between inputs that satisfy  $P$  and inputs that are  $\epsilon$ -far from satisfying  $P$  (that is inputs whose farness from any object from  $P$  is more than  $\epsilon$ ) while using at most  $q$  queries. By their nature the only possible testing algorithms are probabilistic, with either 1-sided or 2-sided error (1-sided error algorithms must accept objects from  $P$  with probability 1). Traditionally the query “what is  $|w|$ ” is not counted towards the  $q$  query limit.

The ultimate goal of Property-Testing research is to classify properties according to their optimal  $\epsilon$ -test query-complexity. In particular, a property whose optimal query complexity depends on  $\epsilon$

---

\*Research supported in part by an ERC-2007-StG grant number 202405.

<sup>†</sup>Department of Computer Science, Technion, Haifa 32000, Israel. eldar@cs.technion.ac.il

<sup>‡</sup>Department of Computer Science, Technion, Haifa 32000, Israel. jongold@cs.technion.ac.il

<sup>§</sup>Birkbeck, University of London, London, UK. oded@dc.s.bbk.ac.uk

alone and not on the length  $|w|$  is called *testable*. In many (but not all) cases a “query-efficient” property test will also be efficient in other computational resources, such as running time (usually it will be the time it takes to retrieve a query multiplied by some function of the number of queries) and space complexity (outside the space used to store the input itself).

Property-Testing was first addressed by Blum, Luby and Rubinfeld [4], and most of its general notions were first formulated by Rubinfeld and Sudan [18], where the investigated properties are mostly of an algebraic nature, such as the property of a Boolean function being linear. The first excursion to combinatorial properties and the formal definition of testability were by Goldreich, Goldwasser and Ron [11]. Since then Property-Testing has attracted significant attention leading to many results. For surveys see [6], [10], [16], [17].

Many times *families* of properties are investigated rather than individual properties, and one way to express such families is through the use of parameters. For example,  $k$ -colorability (as investigated in [11]) has an integer parameter, and the more general partition properties investigated there have the sequence of density constraints as parameters. In early investigations the parameters were considered “constant” with regards to the query complexity bounds, which were allowed to depend on them arbitrarily. However, later investigations involved properties whose “parameter” has in fact a description size comparable to the input itself. Probably the earliest example of this is [14], where properties accepted by a general read-once oblivious branching program are investigated. In such a setting a general dependency on the parameter is inadmissible, and indeed in [14] the dependency is only on the maximum width of the branching program, which may be thought of as a complexity parameter of the stated problem.

A fitting name for such families of properties is *massively parametrized properties*. A good way to formalize this setting is to consider an input to be divided to two parts. One part is the *parameter*, the branching program in the example above, to which the testing algorithm is allowed full access without counting queries. The other part is the *tested input*, to which the algorithm is allowed only a limited number of queries as above. Also, in the definition of fairness only changes to the tested input are allowed, and not to the parameter. In other words, two “inputs” that differ on the parameter part are considered to be  $\infty$ -far. In this setting also other computational measures commonly come into play, such as the running time it takes to plan which queries will be made to the tested input.

Recently, a number of results concerning a massively parametrized setting (though at first not under this name) have appeared. See for example [12, 5, 7, 9] and the survey [15], as well as [2], where such an  $\epsilon$ -test was used as part of a larger mechanism.

A central area of research in Property-Testing in general and Massively-Parametrized Testing in particular is to associate the query complexity of problems to their other measures of complexity. There are a number of results in this direction, to name some examples see [1, 14, 8]. In [3] the study of formulae satisfiability was initiated. There it was shown that there exists a property that is defined by a 3-CNF formula and yet has a query complexity that is linear in the size of the input. This implies that knowing that a specific property is accepted by a 3-CNF formula does not give any information about its query complexity. In [13] it was shown that if a property is accepted by a read-twice CNF formula, then the property is testable. Here we continue this line of research.

In this paper we study the query complexity of properties that are accepted by read once formulae. These can be described as computational trees, with the tested input values at the leaves

and logic gates at the other nodes, where for an input to be in the property a certain value must result when the calculation is concluded at the root.

We prove a number of results. Section 2 contains preliminaries. First we define the properties we test, and then we introduce numerous definitions and lemmas about bringing the formulas whose satisfaction is tested into a normalized "basic form". These are important and in fact implicitly form a preprocessing part of our algorithms. Once the formula is put in a basic form, testing an assignment to the formula becomes manageable.

In Section 3 we show the testability of properties defined by formulae involving arbitrary gates of bounded arity. For such formula involving only monotone gates, we provide an *estimation* algorithm in Section 4, that is an algorithm that not only tests for the property but with high probability outputs a real number  $\eta$  such that the true farness of the tested input from the property is between  $\eta - \epsilon$  and  $\eta + \epsilon$ . In Section 5 we show that when restricted to And/Or gates, we can provide a test whose query complexity is quasipolynomial in  $\epsilon$ .

On the other hand, we prove in Section 7 of in the appendix that these results can not be generalized to alphabets that have at least four different letters. We construct a formula utilizing only one (symmetric and binary) gate type over an alphabet of size 4, such that the resulting property requires a number of queries depending on the formula (and input) size for a  $1/4$ -test.

We supply a brief analysis of the running times of the algorithms in Section 6. One interesting implication of the testability results presented here, is that any read-once formula accepting an untestable Boolean property must use unbounded arity gates other than And/Or.

## Acknowledgment

We thank Prajakta Nimbhorkar for the helpful discussion during the early stages of this work.

## 2 Preliminaries

We use  $[k]$  to denote the set  $\{1, \dots, k\}$ . A *digraph*  $G$  is a pair  $(V, E)$  such that  $E \subseteq V \times V$ . For every  $v \in V$  we set  $\text{out-deg}(v) = \{u \in V \mid (u, v) \in E\}$ . A *path* is a tuple  $(u_1, \dots, u_k) \in |V|^k$  such that  $u_1, \dots, u_k$  are all distinct and  $(u_i, u_{i+1}) \in E$  for every  $i \in [k-1]$ . The *length* of a path  $(u_1, \dots, u_k) \in |V|^k$  is  $k-1$ . We say that there is a path from  $u$  to  $v$  if there exists a path  $(u_1, \dots, u_k)$  in  $G$  such that  $u_1 = u$ , and  $u_k = v$ . The *distance* from  $u \in V$  to  $v \in V$ , denoted  $\text{dist}(u, v)$ , is the length of the shortest path from  $u$  to  $v$  if one exists and infinity otherwise.

We use the standard terminology for outward-directed rooted trees. A *rooted directed tree* is a tuple  $(V, E, r)$ , where  $(V, E)$  is a digraph,  $r \in V$  and for every  $v \in V$  there is a unique path from  $r$  to  $v$ . Let  $u, v \in V$ . If  $\text{out-deg}(v) = 0$  then we call  $v$  a leaf. We say that  $u$  is an *ancestor* of  $v$  and  $v$  is a *descendant* of  $u$  if there is a path from  $u$  to  $v$ . We say that  $u$  is a *child* of  $v$  and  $v$  is a *parent* of  $u$  if  $(v, u) \in E$ , and set  $\text{Children}(v) = \{w \in V \mid w \text{ is a child of } v\}$ .

### 2.1 Formulae, evaluations and testing

With the terminology of rooted trees we now define our properties; first we define what is a formula and then we define what it means to satisfy one.

**Definition 2.1 (Formula)** A Read-Once Formula is a tuple  $\Phi = (V, E, r, X, \kappa, B, \Sigma)$ , where  $(V, E, r)$  is a rooted directed tree,  $\Sigma$  is an alphabet,  $X$  is a set of variables (later on they will take values in  $\Sigma$ ),  $B \subseteq \bigcup_{k < \infty} \{\Sigma^k \mapsto \Sigma\}$  a set of functions over  $\Sigma$ , and  $\kappa : V \rightarrow B \cup X \cup \Sigma$  satisfies the following (we abuse notation somewhat by writing  $\kappa_v$  for  $\kappa(v)$ ).

- For every leaf  $v \in V$  we have that  $\kappa_v \in X \cup \Sigma$ .
- For every  $v$  that is not a leaf  $\kappa_v \in B$  is a function whose arity is  $|\text{Children}(v)|$ .

In the case where  $B$  contains functions that are not symmetric, we additionally assume that for every  $v \in V$  there is an ordering of  $\text{Children}(v) = (u_1, \dots, u_k)$ .

In the special case where  $\Sigma$  is the binary alphabet  $\{0, 1\}$ , we say that  $\Phi$  is *Boolean*. Unless stated otherwise  $\Sigma = \{0, 1\}$ , in which case we shall omit  $\Sigma$  from the definition of formulae. A formula  $\Phi = (V, E, r, X, \kappa, B, \Sigma)$  is called *read  $k$ -times* if for every  $x \in X$  there are at most  $k$  vertices  $v \in V$ , where  $\kappa_v \equiv x$ . We call  $\Phi$  a *read-once-formula* if it is read 1-times. A formula  $\Phi = (V, E, r, X, \kappa, B, \Sigma)$  is called  *$k$ -ary* if the arity (number of children) of all its vertices is at most  $k$ . If a formula is 2-ary we then call it *binary*. A function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  is *monotone* if whenever  $x \in \{0, 1\}^n$  is such that  $f(x) = 1$ , then for every  $y \in \{0, 1\}^n$  such that  $x \leq y$  (coordinate-wise) we have  $f(y) = 1$  as well. If all the functions in  $B$  are monotone then we say that  $\Phi$  is (explicitly) *monotone*. We denote  $|\Phi| = |X|$  and call it the *formula size*.

**Definition 2.2 (Sub-Formula)** Let  $\Phi = (V, E, r, X, \kappa, B)$  be a formula and  $u \in V$ . The formula  $\Phi_u = (V_u, E_u, u, X_u, \kappa, B)$ , is such that  $V_u \subseteq V$ , with  $v \in V_u$  if and only if  $\text{dist}(u, v)$  is finite, and  $(v, w) \in E_u$  if and only if  $v, w \in V_u$  and  $(v, w) \in E$ .  $X_u$  is the set of all  $\kappa_v \in X$  such that  $v \in V_u$ . If  $u \neq r$  then we call  $\Phi_u$  a *strict sub-formula*. We define  $|\Phi_u|$  to be the number of variables in  $V_u$ , that is  $|\Phi_u| = |X_u|$ .

**Definition 2.3 (assignment to and evaluation of a formula)** An assignment  $\sigma$  to a formula  $\Phi = (V, E, r, X, \kappa, B, \Sigma)$  is a mapping from  $X$  to  $\Sigma$ . The evaluation of  $\Phi$  given  $\sigma$ , denoted (abusing notation somewhat) by  $\sigma(\Phi)$ , is defined as  $\sigma(r)$  where  $\sigma : V \rightarrow \Sigma$  is recursively defined as follows.

- If  $\kappa_v \in \Sigma$  then  $\sigma(v) = \kappa_v$ .
- If  $\kappa_v \in X$  then  $\sigma(v) = \sigma(\kappa_v)$ .
- Otherwise ( $\kappa_v \in B$ ) we set  $\sigma(v) = \kappa_v(\sigma(u_1), \dots, \sigma(u_k))$ , where  $\text{Children}(v) = (u_1, \dots, u_k)$ .

Given an assignment  $\sigma : X \rightarrow \Sigma$  and  $u \in V$ , we let  $\sigma_u$  denote its restriction to  $X_u$ , but whenever there is no confusion we just use  $\sigma$  also for the restriction (as an assignment to  $\Phi_u$ ).

For Boolean formulae, we set  $\text{SAT}(\Phi = b)$  to be all the assignments  $\sigma$  to  $\Phi$  such that  $\sigma(\Phi) = b$ . When  $b = 1$  and we do not consider the case  $b = 0$  in that context, then we simply denote these assignments by  $\text{SAT}(\Phi)$ . If  $\sigma \in \text{SAT}(\Phi)$  then we say that  $\sigma$  *satisfies*  $\Phi$ . Let  $\sigma_1, \sigma_2$  be assignments to  $\Phi$ . We define  $\text{farness}_\Phi(\sigma_1, \sigma_2)$  to be the relative Hamming distance between the two assignments. That is,  $\text{farness}_\Phi(\sigma_1, \sigma_2) = |\{x \in X \mid \sigma_1(x) \neq \sigma_2(x)\}|/|\Phi|$ . For every subset  $S$  of assignments to  $\Phi$  we set  $\text{farness}_\Phi(\sigma, S) = \min\{\text{farness}_\Phi(\sigma, \sigma') \mid \sigma' \in S\}$ . If  $\text{farness}_\Phi(\sigma, S) > \epsilon$  then  $\sigma$  is  $\epsilon$ -far from  $S$  and otherwise it is  $\epsilon$ -close to  $S$ .

We now have the ingredients to define testing of assignments to formulae in a massively parametrized model. Namely, the formula  $\Phi$  is the parameter that is known to the algorithm in advance and may not change, while the assignment  $\sigma : X \rightarrow \Sigma$  must be queried with as few queries as possible, and fairness is measured with respect to the fraction of alterations it requires.

**Definition 2.4**  $[(\epsilon, q)\text{-test}]$  *An  $(\epsilon, q)$ -test for  $SAT(\Phi)$  is a randomized algorithm  $\mathcal{A}$  with free access to  $\Phi$ , that given oracle access to an assignment  $\sigma$  to  $\Phi$  operates as follows.*

- $\mathcal{A}$  makes at most  $q$  queries to  $\sigma$  (where on a query  $x \in X$  it receives  $\sigma_x$  as the answer).
- If  $\sigma \in SAT(\Phi)$ , then  $\mathcal{A}$  accepts (returns 1) with probability at least  $2/3$ .
- If  $\sigma$  is  $\epsilon$ -far from  $SAT(\Phi)$ , then  $\mathcal{A}$  rejects (returns 0) with probability at least  $2/3$ . Recall that  $\sigma$  is  $\epsilon$ -far from  $SAT(\Phi)$  if its relative Hamming distance from every assignment in  $SAT(\Phi)$  is at least  $\epsilon$ .

We say that  $\mathcal{A}$  is non-adaptive if its choice of queries is independent of their values. We say that  $\mathcal{A}$  has 1-sided error if given oracle access to  $\sigma \in SAT(\Phi)$ , it accepts (returns 1) with probability 1. We say that  $\mathcal{A}$  is an  $(\epsilon, q)$ -estimator if it returns a value  $\eta$  such that with probability at least  $2/3$ ,  $\sigma$  is both  $\eta + \epsilon$ -close and  $\eta - \epsilon$ -far from  $SAT(\Phi)$ .

We can now summarize the contributions of the paper in the following theorem:

**Theorem 2.5 (Main Theorem)** *The following theorems all hold:*

- For any read-once formula  $\Phi$  where  $B$  is the set of all functions of arity at most  $k$  there exists a 1-sided  $(\epsilon, q)$ -test for  $SAT(\Phi)$  with  $q = \exp(\text{poly}(\epsilon^{-1}))$ .
- For any read-once formula  $\Phi$  where  $B$  is the set of all monotone functions of arity at most  $k$  there exists an  $(\epsilon, q)$ -estimator for  $SAT(\Phi)$  with  $q = \exp(\text{poly}(\epsilon^{-1}))$ .
- For any read-once formula  $\Phi$  where  $B$  is the set of all conjunctions and disjunctions of any arity there exists an  $(\epsilon, q)$ -test for  $SAT(\Phi)$  with  $q = \epsilon^{O(\log \epsilon)}$ .
- There exists an infinite family of 4 valued read-once formulae  $\Phi$  such that there is no non-adaptive  $(\epsilon, q)$ -test for  $SAT(\Phi)$  with  $q = O(\text{depth}(\Phi))$ , and no adaptive  $(\epsilon, q)$ -test for  $SAT(\Phi)$  with  $q = O(\log(\text{depth}(\Phi)))$ .

Note that for the first two items, the degree of the polynomial is linear in  $k$ .

## 2.2 Basic formula simplification and handling

In the following, unless stated otherwise, our formulae will all be read-once and Boolean. For our algorithms to work, we will need a somewhat “canonical” form of such formulae. We say that two formulae  $\Phi$  and  $\Phi'$  are *equivalent* if  $\sigma(\Phi) = \sigma(\Phi')$  for every assignment  $\sigma : X \rightarrow \Sigma$ .

**Definition 2.6** *The mDNF (monotone disjunctive normal form) of a monotone boolean function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  is a set of terms  $T$  where each term  $T_i \in T$  is a subset  $T_i \subseteq [n]$ , there exists no two different terms  $T_i, T_j \in T$  such that  $T_i \subset T_j$ , and for every  $x \in \{0, 1\}^n$ ,  $f(x) = 1$  if and only if there exists a term  $T_j \in T$  such that for all  $i \in T_j$ , we have that  $x_i = 1$ .*

**Observation 2.7** Any monotone boolean function  $f : \{0,1\}^n \rightarrow \{0,1\}$  has a unique mDNF  $T$ .

**Proof.** Create terms for all of the 1-witnesses of the function  $f$ , and proceed to remove all terms  $T_i \in T$  such that there exists a term  $T_j \in T$  and  $T_j \subseteq T_i$ . One can see that this process results in a unique mDNF and that conversely any mDNF must result from this process. ■

**Definition 2.8** For  $u \in V$ ,  $v \in \text{Children}(u)$  is called (a,b)-forceful if  $\sigma(v) = a$  implies  $\sigma(u) = b$ .  $v$  is forceful if it is (a,b)-forceful for some  $a, b \in \{0,1\}$ .

Forceful variables are variables that cause “Or-like” or “And-like” behavior in the gate.

**Definition 2.9** A vertex  $v \in V$  is called unforceable if no child of  $v$  is forceful.

**Definition 2.10 ( $k$ -x-Basic formula)** A read-once formula  $\Phi$  is  $k$ -x-basic if it is Boolean, all the functions in  $B$  have arity at least 2, and are either of arity at most  $k$  and unforceable, or  $\wedge$  or  $\vee$  of arity at least 2, and  $\Phi$  satisfies the following. There is no  $v \in V$  such that  $\kappa_v \in \{0,1\}$ . No  $\wedge$  is a child of a  $\wedge$  and no  $\vee$  is a child of a  $\vee$ . Any variable may appear at most once in a leaf, either positively or negated.

The set of variables that appear negated will be denoted by  $\neg X$ .

**Definition 2.11 ( $k$ -Basic formula)** A read-once formula  $\Phi$  is a  $k$ -basic formula if it is  $k$ -x-basic and furthermore, all unforceable functions in  $B$  are also monotone. If  $B$  contains only conjunctions and disjunctions we abbreviate and call the formula basic.

**Lemma 2.12** Every read-once formula  $\Phi$  with gates of arity at most  $k$  has an equivalent  $k$ -x-basic formula  $\Phi'$ .

**Proof.** Suppose for some  $u$  that  $v \in \text{Children}(u)$  is (a,b)-forceful. If  $b = 1$  then  $\kappa_u$  can be replaced with an  $\vee$  gate, where one input of the  $\vee$  gate is  $v$  if  $a = 1$  or the negation of  $v$  if  $a = 0$ , and the other input is the result of  $u$  when fixing  $\sigma(\kappa_v) = 1 - a$ . If  $b = 0$  then  $\kappa_u$  can be replaced with an  $\wedge$  gate, where one input of the  $\wedge$  gate is  $v$  if  $a = 0$  or the negation of  $v$  if  $a = 1$ , and the other input is the negation of the gate  $u$  when it is assumed that  $\sigma(\kappa_v) = a$ . After performing this transformation sufficiently many times we have no forceable gates left.

We will now eliminate  $\neg$  gates. Any  $\neg$  gate in the input or output of a gate which is not  $\wedge$  or  $\vee$  can be assimilated into the gate. Otherwise, a  $\neg$  on the output of an  $\vee$  can be replaced with an  $\wedge$  with  $\neg$ 's on all of its inputs, according to De-Morgan's laws. Also by De-Morgan's laws, a  $\neg$  on the output of a  $\wedge$  can be replaced with an  $\vee$  with  $\neg$ 's on all of its inputs.

Finally, any  $\vee$  gates that have  $\vee$  children can be merged with them, and the same goes for  $\wedge$  gates. Now we have achieved an equivalent  $k$ -x-basic formula. ■

Note that  $\vee$  and  $\wedge$  gates are very much forceable.

**Observation 2.13** Any formula  $\Phi$  which is comprised of only monotone  $k$ -arity gates has an equivalent  $k$ -basic formula  $\Phi'$ .

### 2.3 Observations about subformulae and farness

**Definition 2.14 (heaviest child  $h(v)$ )** Let  $\Phi = (V, E, r, X, \kappa, B)$  be a formula. For every  $v \in V$  we define  $h(v)$  to be  $v$  if  $\text{Children}(v) = \emptyset$ , and otherwise to be an arbitrarily selected vertex  $u \in \text{Children}(v)$ , such that  $|\Phi_u| = \max\{|\Phi_w| \mid w \in \text{Children}(v)\}$ .

**Definition 2.15 (vertex depth  $\text{depth}_\Phi(v)$ )** Let  $\Phi = (V, E, r, X, \kappa, B)$  be a formula. For every  $v \in V$  we define  $\text{depth}_\Phi(v) = \text{dist}(r, v)$  and  $\text{depth}(\Phi) = \max\{\text{depth}_\Phi(u) \mid u \in V\}$ .

**Observation 2.16** Let  $v \in V$  be such that either  $\kappa_v \equiv \vee$  and  $b = 0$  or  $\kappa_v \equiv \wedge$  and  $b = 1$ , and  $\text{farness}(\sigma, \text{SAT}(\Phi_v = b)) \geq \epsilon$ . For every  $1 > \alpha > 0$  there exists  $S \subseteq \text{Children}(v)$  such that  $\sum_{s \in S} |\Phi_s| \geq \epsilon \alpha^2 |\Phi|$  and  $\text{farness}(\sigma, \text{SAT}(\Phi_w = b)) \geq \epsilon(1 - \alpha)$  for every  $w \in S$ . Furthermore, the exists a child  $u \in \text{Children}(v)$  such that  $\text{farness}(\sigma, \text{SAT}(\Phi_u = b)) \geq \epsilon$ .

**Proof.** Let  $T$  be the maximum subset of  $\text{Children}(v)$  such that  $\Phi_w$  is  $\epsilon(1 - \alpha)$ -far from being evaluated to  $b$  for every  $w \in T$ . If  $\sum_{t \in T} |\Phi_t| < \epsilon \alpha^2 |\Phi|$  then the distance from having  $\Phi_v$  evaluate to  $b$  is at most  $\epsilon \alpha^2 + \epsilon(1 - \alpha)(1 - \alpha) < \epsilon$ , which contradicts the assumption.

For the last part, note that if no such child exists then  $\Phi_v$  is  $\epsilon$ -close to being evaluated to  $b$ . ■

**Observation 2.17** Let  $v \in V$  be such that either  $\kappa_v \equiv \vee$  and  $b = 1$  or  $\kappa_v \equiv \wedge$  and  $b = 0$ , and  $\text{farness}(\sigma, \text{SAT}(\Phi_v = b)) \geq \epsilon$ . For every child  $u \in \text{Children}(v)$ ,  $|\Phi_u| \geq |\Phi|\epsilon$  and  $\text{farness}(\sigma, \text{SAT}(\Phi_u = b)) \geq \epsilon(1 + \epsilon)$ . Furthermore,  $\epsilon \leq 1/2$ , and for any  $u \in \text{Children}(v) \setminus \{h(v)\}$ ,  $\text{farness}(\sigma, \text{SAT}(\Phi_u = b)) \geq 2\epsilon$ .

**Proof.** First suppose that the weight of some child  $u$  is less than  $\epsilon$ . In this case setting  $u$  to  $b$  makes the formula  $\Phi_v$  evaluate to  $b$  by changing less than an  $\epsilon$  fraction of inputs, a contradiction.

Since there are at least two children, every child  $u$  is of weight at most  $1 - \epsilon$  and since setting it to  $b$  would make  $\Phi_v$  evaluate to  $b$ , it is at least  $\epsilon(1 + \epsilon)$ -far from being evaluated to  $b$ .

For the last part, note that since  $|\text{Children}(v)| > 1$ , there exists  $u \in \text{Children}(v)$  such that  $|\Phi_u| \leq |\Phi_v|/2$ . Thus every assignment to  $\Phi_v$  is  $1/2$ -close to an assignment  $\sigma'$  by which  $\Phi_v$  evaluates to  $b$ . Also note that any  $u \in \text{Children}(v) \setminus \{h(v)\}$  is of weight at most  $1/2$ , and therefore if  $\Phi_u$  were  $2\epsilon$ -close to being evaluated to  $b$ ,  $\Phi_v$  was  $\epsilon$ -close to being evaluated to  $b$ . ■

### 2.4 Heavy and Light Children in General Gates

**Definition 2.18** Given a formula  $\Phi$ , a parameter  $\epsilon$  and a vertex  $u$ , we let  $\ell = \ell(u, \epsilon)$  be the smallest integer such that the size of the  $\ell$ 'th largest child of  $u$  is less than  $|\Phi|(4k/\epsilon)^{-\ell}$  if it exists, and set  $\ell = k + 1$  otherwise. The heavy children of  $u$  are the  $\ell - 1$  largest children of  $u$ , and the rest of the children of  $u$  are its light children.

**Lemma 2.19** If an unforceable vertex  $v$  has a child  $u$  such that  $|\Phi_v|(1 - \epsilon) \leq |\Phi_u|$ , then  $\sigma$  is both  $\epsilon$ -close to  $\text{SAT}(\Phi_v = 1)$  and  $\epsilon$ -close to  $\text{SAT}(\Phi_v = 0)$ .

**Proof.** The child is unforceful, and therefore it is possible to change the remaining children to obtain any output value. ■

**Observation 2.20** *If  $\kappa_u \not\equiv \wedge$  and  $\kappa_u \notin X$  and  $\sigma$  is  $\epsilon$ -far from  $\text{SAT}(\Phi_u = b)$ , then it must have at least two heavy children.*

**Proof.** By the definition of  $\ell$ , if there is just one heavy child, then  $\ell = 2$  and the total weight of the light children is strictly smaller than  $\epsilon$ . Therefore by Lemma 2.19 there must be more than one heavy child, as otherwise the gate is  $\epsilon$ -close to both 0 and 1. ■

### 3 Upper Bound for General Bounded Arity Formula

Algorithm 1 tests whether the input is  $\epsilon$ -close to having output  $b$  with 1-sided error, and also receives a confidence parameter  $\delta$ . The explicit confidence parameter makes the inductive arguments easier and clearer.

---

#### Algorithm 1

---

**Input:** read-once  $k$ -x-basic formula  $\Phi = (V, E, r, X, \kappa)$ , parameters  $\epsilon, \delta > 0, b \in \{0, 1\}$ , oracle to  $\sigma$ .

**Output:** “true” or “false”.

```

1: if  $\epsilon > 1$  then return “true”
2: if  $\kappa_r \in X$  then return the truth value of  $\sigma(r) = b$ 
3: if  $\kappa_r \in \neg X$  then return the truth value of  $\sigma(r) = 1 - b$ 
4: if  $(\kappa_r \equiv \wedge \text{ and } b = 1)$  or  $(\kappa_r \equiv \vee \text{ and } b = 0)$  then
5:    $y \leftarrow \text{“true”}$ 
6:   for  $i = 1$  to  $l = 32(2k/\epsilon)^{2k} \log(\delta^{-1})$  do
7:      $u \leftarrow$  a vertex in  $\text{Children}(r)$  selected independently at random, where the probability
       that  $w \in \text{Children}(r)$  is selected is  $|\Phi_w|/|\Phi|$ 
8:      $y \leftarrow y \wedge \text{Algorithm 1}(\Phi_u, (\epsilon(1 - (2k/\epsilon)^{-k}/16)), \sigma, \delta/2, b)$ 
9:   end for
10:  return  $y$ 
11: end if
12: if  $(\kappa_r \equiv \wedge \text{ and } b = 0)$  or  $(\kappa_r \equiv \vee \text{ and } b = 1)$  then
13:   if there exists a child of weight less than  $\epsilon$  then return “true”
14:    $y \leftarrow \text{“false”}$ 
15:   for all  $u \in \text{Children}(r)$  do  $y \leftarrow y \vee \text{Algorithm 1}(\Phi_u, (\epsilon(1 + \epsilon)), \sigma, \epsilon\delta/2, b)$ 
16:   return  $y$ 
17: end if
18: if there is a child of weight at least  $1 - \epsilon$  then return “true”
19: for all  $u \in \text{Children}(r)$  do
20:    $y_u^0 \leftarrow \text{Algorithm 1}(\Phi_u, (\epsilon(1 + (4k/\epsilon)^{-k})), \sigma, \delta/2k, 0)$ 
21:    $y_u^1 \leftarrow \text{Algorithm 1}(\Phi_u, (\epsilon(1 + (4k/\epsilon)^{-k})), \sigma, \delta/2k, 1)$ 
22: end for
23: if There exists a string  $x \in \{0, 1\}^k$  such that  $\kappa_r$  on  $x$  would evaluate to  $b$  and for all  $u \in$ 
     $\text{Children}(r)$  we have  $y_u^{x_u}$  equal to “true” then return “true” else return “false”

```

---

**Lemma 3.1** *The depth of recursion in Algorithm 1 is at most  $16(4k/\epsilon)^k \log(\epsilon^{-1})$ .*



**Proof.** If  $\epsilon > 1$  then the condition in Line 1 is satisfied and the algorithm returns without making any queries.

All recursive calls occur in Lines 8, 15, 20 and 21.

Since  $\Phi$  is  $k$ - $x$ -basic, any call with a subformula whose root is labeled by  $\wedge$  results in calls to subformulas, each with a root labeled either by  $\vee$  or an unforceable gate, and with the same  $b$  value (this is crucial since the  $b$  value for which  $\wedge$  recurses with a smaller  $\epsilon$  is the  $b$  value for which  $\vee$  recurses with a bigger  $\epsilon$ , and vice-versa). Similarly, any call with a subformula whose root is labeled by  $\vee$  results in calls to subformulas, each with a root labeled either by  $\wedge$  or an unforceable gate, and with the same  $b$  value. Therefore, an increase of two in the depth results in an increase of the fairness parameter from  $\epsilon$  to at least  $(\epsilon(1 - (2k/\epsilon)^{-k}/16))(\epsilon(1 + (4k/\epsilon)^{-k})) \geq \epsilon(1 + (4k/\epsilon)^{-k}/16)$ . Thus in recursive calls of depth  $16(4k/\epsilon)^k \log(\epsilon^{-1})$  the fairness parameter exceeds 1 and the call returns without making any further calls. ■

**Lemma 3.2** *Algorithm 1 uses at most  $\epsilon^{-480(4k/\epsilon)^{k+3} \log \log(\delta^{-1})}$  queries.*

**Proof.** If  $\epsilon > 1$  then the condition in Line 1 is satisfied and no queries are made. Therefore assume  $\epsilon \leq 1$ . Observe that in a specific instantiation at most one query is used, either in Line 2 or Line 3. Therefore the number of queries is upper bounded by the number of instantiations of Algorithm 1.

In a specific instantiation at most  $32(2k/\epsilon)^{2k} \log(\delta^{-1})$  recursive calls are made in total (note that by Line 13 there are at most  $1/\epsilon$  children in the case of the condition in Line 12, and in the case of an unforceable gate there are at most  $2k$  recursive calls). Recall that by Lemma 3.1 the depth of the recursion is at most  $16(4k/\epsilon)^k \log(\epsilon^{-1})$ .

To conclude, we note that the value of the confidence parameter in all these calls is lower bounded by  $\delta \cdot (\epsilon/2k)^{16(4k/\epsilon)^k \log(\epsilon^{-1})} \geq \delta \cdot \epsilon^{32(4k/\epsilon)^k \log(k\epsilon^{-1})}$ . Therefore at most  $(32(4k/\epsilon)^{2k} \log(\delta \cdot \epsilon^{-32(2k/\epsilon)^k \log(k\epsilon^{-1})}))^{16(4k/\epsilon)^k \log(\epsilon^{-1})} = \epsilon^{-480(4k/\epsilon)^{k+3} \log \log(\delta^{-1})}$  queries are used. ■

**Lemma 3.3** *If  $\Phi$  on  $\sigma$  evaluates to  $b$  then Algorithm 1 returns “true” with probability 1.*

**Proof.** If  $\epsilon > 1$  then the condition of Line 1 is satisfied and “true” is returned correctly. We proceed with induction over the depth of the formula. If  $\text{depth}(\Phi) = 0$  then  $\kappa_r \in X \cup \neg X$ . If  $\kappa_r \in X$  then since  $\Phi$  evaluates to  $b$ ,  $\sigma(r) = b$ , if  $\kappa_r \in \neg X$  then  $\sigma(r) = 1 - b$ , and the algorithm returns “true” correctly.

Now assume that  $\text{depth}(\Phi) > 0$ . Obviously, for all  $u \in \text{Children}(r)$ ,  $\text{depth}(\Phi) > \text{depth}(\Phi_u)$  and therefore from the induction hypothesis any recursive call on a subformula that evaluates to  $b'$  returns “true” with probability 1.

If  $\kappa_r \equiv \wedge$  and  $b = 1$  or  $\kappa_r \equiv \vee$  and  $b = 0$ , then it must be the case that for all  $u \in \text{Children}(r)$ ,  $\Phi_u$  evaluates to  $b$ . By the induction hypothesis all recursive calls will return “true” and  $y$  will get the value “true”, which will be returned by the algorithm.

Now assume that  $\kappa_r \equiv \wedge$  and  $b = 0$  or  $\kappa_r \equiv \vee$  and  $b = 1$ . Since  $\Phi$  evaluates to  $b$  then it must be the case that at least for one  $u \in \text{Children}(r)$ ,  $\Phi_u$  evaluates to  $b$ . By the induction hypothesis, the recursive call on that  $u$  will return “true”, and  $y$  will get the value “true” which will be returned by the algorithm (unless the algorithm already returned “true” for another reason).

Lastly, assume that  $r$  is an unforceable gate. Since  $\Phi$  evaluates to  $b$ , the children of  $r$  evaluate to an assignment  $x$  to  $\kappa_r$  which evaluates to  $b$ . By the induction hypothesis, for every  $u \in \text{Children}(r)$

the recursive call on  $\Phi_u$  with  $x_u$  will return “true”, and thus the assignment  $x$  will, in particular, fill the condition in Line 23 and the algorithm will return “true”. ■

**Lemma 3.4** *If  $\sigma$  is  $\epsilon$ -far from getting  $\Phi$  to output  $b$  then Algorithm 1 returns “false” with probability at least  $1 - \delta$ .*

**Proof.** The proof is by induction over the tree structure, where we partition to cases according to  $\kappa_r$  and  $b$ . Note that  $\epsilon \leq 1$

If  $\kappa_r \in X$  or  $\kappa_r \in \neg X$  then by Lines 2 or 3 the algorithm returns “false” if  $\sigma$  is 0-far from getting  $\Phi$  to output  $b$ .

If  $\kappa_r \equiv \wedge$  and  $b = 1$  or  $\kappa_r \equiv \vee$  and  $b = 0$ , since  $\sigma$  is  $\epsilon$ -far from getting  $\Phi$  to output  $b$  then by Observation 2.16 we get that there exists  $T \subseteq \text{Children}(r)$  such that  $\sum_{t \in T} |\Phi_t| \geq |\Phi| \epsilon ((2k/\epsilon)^{-2k}/16)$  and each  $\Phi_t$  is  $\epsilon(1 - (2k/\epsilon)^{-k}/16)$ -far from being evaluated to  $b$ . Let  $S$  be the set of all vertices selected in Line 7. The probability of a vertex from  $T$  being selected is at least  $\epsilon((2k/\epsilon)^{-2k}/16)$ . Since this happens at least  $32(2k/\epsilon)^{2k} \log(\delta^{-1})$  times independently, with probability at least  $1 - \delta/2$  we have that  $S \cap T \neq \emptyset$ . Letting  $w \in T \cap S$ , the recursive call on it with parameter  $\epsilon(1 - (2k/\epsilon)^{-k}/16)$  will return “false” with probability at least  $1 - \delta/2$ , which will eventually cause the returned value to be “false” as required. Thus the algorithm succeeds with probability at least  $1 - \delta$ .

Now assume that  $\kappa_r \equiv \wedge$  and  $b = 0$  or  $\kappa_r \equiv \vee$  and  $b = 1$ . Since  $\Phi$  is  $\epsilon$ -far from being evaluated to  $b$ , Observation 2.17 implies that all children are of weight at least  $\epsilon$ , and therefore the conditions of Line 13 would not be triggered. Every recursive call on a vertex  $v \in \text{Children}(r)$  is made with distance parameter  $\epsilon(1 + \epsilon)$  and so it returns “true” with probability at most  $\epsilon\delta/2$ . Since there are at most  $\epsilon^{-1}$  children of  $r$ , the probability that none returns “true” is at least  $1 - \delta/2$  and in that case the algorithm returns “false” successfully.

Now assume that  $\kappa_r$  is some unforceable gate. By Observation 2.19, since  $\Phi$  is  $\epsilon$ -far from being satisfied the condition in Line 18 is not triggered. If the algorithm returned “true” then it must be that the condition in Line 23 is satisfied. If there exists some heavy  $u \in \text{Children}(r)$  such that  $y_u^b$  is “true” and  $y_u^{1-b}$  is “false”, then by Lemma 3.3 the formula  $\Phi_u$  does evaluate to  $b$  and the string in  $x$  must be such that  $x_u = b$ . For the rest of the children of  $r$ , assuming the calls succeeded, each the subformula rooted in  $v$  is  $(\epsilon(1 + (4k/\epsilon)^{-k}))$ -close to evaluate to  $x_v$ . Since  $u$  is heavy, the total weight of  $\text{Children}(r) \setminus \{u\}$  is at most  $1 - (4k/\epsilon)^{-k}$ , and thus by changing at most a  $(\epsilon(1 + (4k/\epsilon)^{-k}))(1 - (4k/\epsilon)^{-k}) \leq \epsilon$  fraction of inputs we can get to an assignment where  $\Phi$  evaluates to  $b$ .

If all heavy children  $u$  are such that both  $y_u^b$  and  $y_u^{1-b}$  are “true”, then pick some heavy child  $u$  arbitrarily. Since  $r$  is unforceable, there is an assignment that evaluates to  $b$  no matter what the value of  $\Phi_u$  is. Take such an assignment  $x$  that fits the real value of  $\Phi_u$ . Note that for every heavy child  $v$  we have that  $y_v^{x_v}$  is “true”, and therefore by changing at most an  $(\epsilon(1 + (4k/\epsilon)^{-k}))$ -fraction of the variables in  $\Phi_v$  we can get it to evaluate to  $x_v$ . The weight of  $u$  is at least  $(4k/\epsilon)^{-\ell+1}$ , thus the total weight of the other heavy children is at most  $1 - (4k/\epsilon)^{-\ell+1}$  and the total weight of the light children is at most  $\frac{\epsilon}{4}(4k/\epsilon)^{-\ell}$ . So by changing all subformulas to evaluate to the value implied by  $x$  we change at most an  $(\epsilon(1 + (4k/\epsilon)^{-k}))(1 - (4k/\epsilon)^{-\ell+1}) + \frac{\epsilon}{4}(4k/\epsilon)^{-\ell} \leq \epsilon$  fraction of inputs and get an assignment where  $\Phi$  evaluates to  $b$ . Note that this  $x$  is not necessarily the one found in Line 23.

Thus we have found that finding an assignment  $x$  in Line 23, assuming the calls are correct, implies that  $\Phi$  is  $\epsilon$ -close to evaluate to  $b$ . The probability that all relevant calls to an assignment return “true” incorrectly is at most the probability that the  $2k$  recursive calls err, which by the union bound is at most  $\delta$ , and the algorithm will return “false” correctly with probability at least  $1 - \delta$ .  $\blacksquare$

## 4 Estimator for monotone formula of bounded arity

Algorithm 2 operates in a recursive manner, and estimates the distance to satisfying the formula rooted in  $r$  according to estimates for the subformula rooted in every children of  $r$ . The algorithm explicitly receives a confidence parameter  $\delta$  as well as the approximation parameter  $\epsilon$ , and should with probability at least  $1 - \delta$  return a number  $\eta$  such that the input is both  $(\eta + \epsilon)$ -close and  $(\eta - \epsilon)$ -far from satisfying the given formula. The explicit confidence parameter makes the inductive arguments easier and clearer.

---

### Algorithm 2

---

**Input:** read-once  $k$ -basic formula  $\Phi = (V, E, r, X, \kappa)$ , parameters  $\epsilon, \delta > 0$ , oracle to  $\sigma$ .

**Output:**  $\eta \in [0, 1]$ .

```

1: if  $\kappa_r \in X$  then return  $1 - \sigma(\kappa_r)$ 
2: if  $\epsilon > 1$  then return 0
3: if  $\kappa_r \equiv \vee$  and there exists  $u \in \text{Children}(r)$  with  $|\Phi_u| < \epsilon|\Phi|$  then return 0
4: if  $\kappa_r \equiv \wedge$  then
5:   for  $i = 1$  to  $l = \lceil 1000\epsilon^{-2k-2}(4k)^{2k} \cdot \log(1/\delta) \rceil$  do
6:      $u \leftarrow$  a vertex in  $\text{Children}(r)$  selected independently at random, where the probability
       that  $w \in \text{Children}(r)$  is selected is  $|\Phi_w|/|\Phi|$ 
7:      $\alpha_i \leftarrow$  Algorithm 2( $\Phi_u, \epsilon(1 - (4k/\epsilon)^{-k}/8), \delta\epsilon(4k/\epsilon)^{-k}/16, \sigma$ )
8:   end for
9:   return  $\sum_{i=1}^l \alpha_i / l$ 
10: else if  $\kappa_r \not\equiv \wedge$  then
11:   for every light child  $u$  of  $r$  set  $\alpha_u \leftarrow 0$ 
12:   for every heavy child  $u$  of  $r$  set  $\alpha_u \leftarrow$  Algorithm 2( $\Phi_u, \epsilon(1 + (4k/\epsilon)^{-k}), \delta/\max\{k, 1/\epsilon\}, \sigma$ )
13:   for every term  $C$  in the mDNF of  $\kappa_r$  set  $\alpha_C \leftarrow \sum_{u \in C} \alpha_u \cdot \frac{|\Phi_u|}{|\Phi|}$ 
14:   return  $\min\{\alpha_C : C \in \text{mDNF}(\kappa_r)\}$ 
15: end if

```

---

The following states that Algorithm 2 indeed gives an estimation of the distance. While estimation algorithms cannot have 1-sided error, there is an additional feature of this algorithm that makes it also useful as a 1-sided test (by running it and accepting if it returns  $\eta = 0$ ).

**Theorem 4.1** *With probability at least  $1 - \delta$ , the output of Algorithm 2( $\Phi, \epsilon, \delta, \sigma$ ) is an  $\eta$  such that the assignment  $\sigma$  is both  $(\eta + \epsilon)$ -close to satisfying  $\Phi$  and  $(\eta - \epsilon)$ -far from satisfying it. Additionally, if the assignment  $\sigma$  satisfies  $\Phi$  then  $\eta = 0$  with probability 1. Its query complexity (treating  $k$  and  $\delta$  as constant) is always  $O(\exp(\text{poly}(\epsilon^{-1})))$ .*

**Proof.** The bound on the number of queries is a direct result of Lemma 4.3 below. Given that, the correctness proof is done by induction on the height of the formula. The base case (for any  $\epsilon$  and  $\delta$ ) is the observation that an instantiation of the algorithm that makes no recursive calls (i.e. triggers the condition in Line 1 or 2) always gives a value that satisfies the assertion.

The induction step uses Lemma 4.4 and Lemma 4.5 below. Given that the algorithm performs correctly (for any  $\epsilon$  and  $\delta$ ) for every formula  $\Phi'$  of height smaller than  $\Phi$ , the assertions of the lemma corresponding to  $\kappa_r$  (out of the two) are satisfied, and so the correctness for  $\Phi$  itself follows. ■

The dependency on  $\delta$  can be made into a simple logarithm by a standard amplification technique: Algorithm 2 is run  $O(1/\delta)$  independent times, each with a confidence parameter  $2/3$ , and then the median of the outputs is taken.

**Lemma 4.2** *When called with  $\Phi$ ,  $\epsilon$ ,  $\delta$ , and oracle access to  $\sigma$ , Algorithm 2 goes down at most  $2(4k/\epsilon)^k \log(1/\epsilon) = \text{poly}(\epsilon)$  recursion levels. In those recursion levels,  $\delta$  decreases by a factor of at most  $(\epsilon(4k/\epsilon)^{-k}/16)^{2(4k/\epsilon)^k \log(1/\epsilon)} = \exp(\text{poly}(1/\epsilon))$ .*

**Proof.** Recursion can only happen on Line 7 and Line 12. Moreover, because of the formula being  $k$ -basic, recursion cannot follow through Line 7 two recursion levels in a row. ■

**Lemma 4.3** *When called with  $\Phi$ ,  $\epsilon$ ,  $\delta$ , and oracle access to  $\sigma$ , Algorithm 2 uses a total of at most  $\exp(\text{poly}(1/\epsilon))$  queries.*

**Proof.** Denoting by  $\delta'$  the smallest  $\delta$  in a recursive call, we have  $\delta' \geq \delta(\epsilon(4k/\epsilon)^{-k}/16)^{2(4k/\epsilon)^k \log(1/\epsilon)}$  by Lemma 4.2. The number of recursive calls per instantiation of the algorithm is thus at most  $l' = \lceil 1000\epsilon^{-2k-2}(4k)^{2k} \cdot \log(1/\delta') \rceil = \text{poly}(1/\epsilon)$ . As the algorithm may make at most one query per instantiation, and this only in the case where a recursive call is not performed, the total number of queries is (bounding the recursion depth through Lemma 4.2) at most  $(l')^{2(4k/\epsilon)^k \log(1/\epsilon)} = \exp(\text{poly}(1/\epsilon))$ . ■

**Lemma 4.4** *If  $\kappa_r \not\equiv \wedge$  and all recursive calls satisfy the assertion of Theorem 4.1, then with probability at least  $1 - \delta$  the current instantiation of Algorithm 2 provides a value  $\eta$  such that  $\sigma$  is both  $(\eta + \epsilon)$ -close to satisfying  $\Phi$  and  $(\eta - \epsilon)$ -far from satisfying it. Furthermore, if  $\sigma$  satisfies  $\Phi$  then with probability 1 the output is  $\eta = 0$ .*

**Proof.** First we note that Step 3, if triggered, gives a correct value for  $\eta$  (as the  $\sigma$  can be made into a satisfying assignment by changing possibly all variables of the smallest child of  $r$ ). We also note that if  $\kappa_r \equiv \vee$  and Step 3 was not triggered, then by definition all of  $r$ 's children are heavy, and there are no more than  $1/\epsilon$  of them.

The true farness of  $\sigma$  from  $\Phi$  is the minimum over all terms  $C$  in  $\kappa_r$  of the adjusted cost of making all children of  $C$  evaluate to 1, which is  $\sum_{u \in C} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|}$ . Now in this case there are clearly no more than  $\max\{k, \epsilon^{-1}\}$  children, and so by the union bound, with probability at least  $1 - \delta$ , every call done through Line 7 gave a value  $\eta_u$  so that indeed  $\sigma$  is  $(\eta_u + \epsilon(1 + (4k/\epsilon)^{-k}))$ -close and  $(\eta_u - \epsilon(1 + (4k/\epsilon)^{-k}))$ -far from  $\Phi_u$ .

Now let  $D_i$  denote  $C_i$  minus any light children that it may contain. It may be that some  $D_i$ 's contain all heavy children, but as there are no forcing children (and there are heavy children)

it must be the case that some  $D_i$ 's do not contain all heavy children, and in Line 14 these will dominate. Note that  $\sum_{u \in D_i} |\Phi_u| \leq (1 - (4k/\epsilon)^{1-\ell})|\Phi|$  for any  $D_i$  not containing a heavy child. This implies by bounding  $(1 + (4k/\epsilon)^{-k}) \cdot (1 - (4k/\epsilon)^{1-\ell})$ :

$$\sum_{u \in D_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} - \epsilon < \frac{\sum_{u \in D_i} \eta_u |\Phi_u|}{|\Phi|} < \sum_{u \in D_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} + \epsilon - 2k(4k/\epsilon)^{-\ell}$$

Now the true farness of  $C_i$  not containing all heavy children is at least that of  $D_i$ , and at most that of  $D_i$  plus with the added farness of making all light children evaluate to 1, which is bounded by  $k(4k/\epsilon)^{-\ell}$ . This means that for such a  $C_i$  we have:

$$\sum_{u \in C_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} - \epsilon < \frac{\sum_{u \in D_i} \eta_u |\Phi_u|}{|\Phi|} < \sum_{u \in C_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} + \epsilon - k(4k/\epsilon)^{-\ell}$$

As the value returned as  $\eta$  is the minimum over terms  $C_i$  in  $\kappa_r$  of  $\eta_u \cdot \frac{\sum_{u \in D_i} |\Phi_u|}{|\Phi|}$ . We also know that this minimum is reached by some  $C_j$  which does not contain all heavy children, but it may be that in fact  $\text{farness}(\sigma, \text{SAT}(\Phi)) = \sum_{u \in C_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|}$  for some  $i \neq j$  (the true farness is the minimum of the total farness of each clause, but it may be reached by a different clause).

By our assumptions

$$\text{farness}(\sigma, \text{SAT}(\Phi)) - \epsilon = \sum_{u \in C_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} - \epsilon \leq \sum_{u \in C_j} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} - \epsilon < \eta$$

so we have one side of the required bound. For the other side, we split into cases. If  $C_i$  also does not contain all heavy children then we use the way we calculated  $\eta$  as the minimum over the corresponding sums:

$$\eta = \frac{\sum_{u \in D_j} \eta_u |\Phi_u|}{|\Phi|} \leq \frac{\sum_{u \in D_i} \eta_u |\Phi_u|}{|\Phi|} < \text{farness}(\sigma, \text{SAT}(\Phi)) + \epsilon$$

In the final case, we note that by the assumptions on the light children we will always have (recalling that  $C_i$  will in particular have all heavy children of  $C_j$ ):

$$\eta = \frac{\sum_{u \in D_j} \eta_u |\Phi_u|}{|\Phi|} < \sum_{u \in C_j} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} + \epsilon - k(4k/\epsilon)^{-\ell} \leq \sum_{u \in C_i} \text{farness}(\sigma, \text{SAT}(\Phi_u)) \cdot \frac{|\Phi_u|}{|\Phi|} + \epsilon$$

where the rightmost term equals  $\text{farness}(\sigma, \text{SAT}(\Phi)) + \epsilon$  as required.

For the last part of the claim, note that if  $\sigma$  satisfies  $\Phi$ , then in particular, one of the terms  $C$  of  $\kappa_r$  must be satisfied. By the induction hypothesis, for all  $u \in C$  we would have  $\alpha_u = 0$  and therefore  $\alpha_C = 0$ , and since  $\alpha$  is taken as a minimum over all terms we would have  $\alpha = 0$ . ■

**Lemma 4.5** *If  $\kappa_r \equiv \wedge$  and all recursive calls satisfy the assertion of Theorem 4.1, then with probability at least  $1 - \delta$  the current instantiation of Algorithm 2 provides a value  $\eta$  such that  $\sigma$  is both  $(\eta + \epsilon)$ -close to satisfying  $\Phi$  and  $(\eta - \epsilon)$ -far from satisfying it. If  $\sigma$  satisfies  $\Phi$  then with probability 1 the output is  $\eta = 0$ .*

**Proof.** First note that if we sample a vertex  $w$  according to the distribution of Line 5 and then take the true farness  $\mathbf{farness}(\sigma, SAT(\Phi_w))$ , then the expectation (but not the value) of this equals  $\mathbf{farness}(\sigma, SAT(\Phi))$ . This is because to make  $\sigma$  evaluate to 1 at the root, we need to make all its children evaluate to 1, an operation whose adjusted cost is given by the weighted sum of farnesses that corresponds to the expectation above.

Thus, denoting by  $X_i$  the random variable whose value is  $\mathbf{farness}(\sigma, SAT(\Phi_{w_i}))$  where  $w_i$  is the vertex picked in the  $i$ th iteration, we have  $E[X_i] = \mathbf{farness}(\sigma, SAT(\Phi))$ . By a Chernoff type bound, with probability at least  $1 - \delta/2$ , the average  $X$  of  $X_1, \dots, X_l$  is no more than  $\epsilon^{k+1}(4k)^{-k}/16$  away from  $E[X_i]$  and hence satisfies:

$$\mathbf{farness}(\sigma, SAT(\Phi)) - \epsilon^{k+1}(4k)^{-k}/16 < X < \mathbf{farness}(\sigma, SAT(\Phi)) + \epsilon^{k+1}(4k)^{-k}/16$$

Then note that by the Markov inequality, the assertion of the lemma means that with probability at least  $1 - \delta/2$ , all calls done in Line 12 but at most  $\epsilon(4k/\epsilon)^{-k}/16$  of them return a value  $\eta_w$  so that  $\sigma$  is  $(\eta_w + \epsilon(1 - (4k/\epsilon)^{-k}/8))$ -close and  $(\eta_w - \epsilon(1 - (4k/\epsilon)^{-k}/8))$ -far from  $\Phi_w$ .

When this happens, at least  $(1 - \epsilon(4k/\epsilon)^{-k}/16)$  of the answers  $\alpha_i$  of the calls are up to  $\epsilon(1 - (4k/\epsilon)^{-k}/16)$  away from each corresponding  $X_i$ , and at most  $\epsilon(4k/\epsilon)^{-k}/16$  of the answers  $\alpha_i$  are up to 1 away from each corresponding  $X_i$ . Summing up these deviations, the final answer average  $\eta$  satisfies

$$X - \epsilon(1 - (4k/\epsilon)^{-k}/4) - \epsilon(4k/\epsilon)^{-k}/16 < \eta < X + \epsilon(1 - (4k/\epsilon)^{-k}/4) + \epsilon(4k/\epsilon)^{-k}/16$$

With probability at least  $1 - \delta$  both of the above events occur, and summing up the two inequalities we obtain the required bound

$$\mathbf{farness}(\sigma, SAT(\Phi)) - \epsilon \leq \eta < \mathbf{farness}(\sigma, SAT(\Phi)) + \epsilon$$

■

## 5 Quasi-polynomial Upper Bound for Basic-Formulae

Let  $\Phi = (V, E, r, X, \kappa, B)$  be a basic formula and  $\sigma$  an assignment to  $\Phi$ .

The main idea of the algorithm is to randomly choose a full root to leaf path, and recurs over all the children of “ $\vee$ ” vertices on this path that go outside of it, if they are not too many. The main technical part is in proving that if  $\sigma$  is indeed  $\epsilon$ -far from satisfying  $\Phi$ , then many of these paths have few such children (few enough to recurs over all of them), where additionally the distance of  $\sigma$  from satisfying the corresponding sub-formulae is significantly larger. An interesting combinatorial corollary of this is that formulae, for which there are not a lot of leaves whose corresponding paths have few such children, do not admit  $\epsilon$ -far assignments at all.

### 5.1 Critical and Important

To understand the intuition behind the following definitions, it is useful to first consider what happens if we could locate a vertex that is “ $(\epsilon, \sigma)$ -critical” in the sense that is defined next.

**Definition 5.1** [  $(\epsilon, \sigma)$ -important,  $(\epsilon, \sigma)$ -critical ] A vertex  $v \in V$  is  $(\epsilon, \sigma)$ -important if  $\sigma \notin \text{SAT}(\Phi)$ , and for every  $u$  that is either  $v$  or an ancestor of  $v$ , we have that

- $\text{farness}(\sigma, \text{SAT}(\Phi_u)) \geq (2\epsilon/3)(1 + 2\epsilon/3)^{\lfloor \text{depth}_\Phi(u)/3 \rfloor}$
- If  $\kappa_u \equiv \vee$  and  $u \neq v$  then  $h(u)$  is either  $v$  or an ancestor of  $v$ .

An  $(\epsilon, \sigma)$ -critical vertex  $v$  is an  $(\epsilon, \sigma)$ -important vertex  $v$  for which  $\kappa_v \in X$ .

Note that such a vertex is never too deep, since  $\text{farness}(\sigma, \text{SAT}(\Phi_u))$  is always at most 1. The following observation follows from Definition 5.1.

**Observation 5.2** If  $v$  is  $(\epsilon, \sigma)$ -important, then  $\text{depth}_\Phi(v) \leq 4\epsilon^{-1} \log(2\epsilon^{-1})$ .

A hypothetical oracle that provides a critical vertex can be used as follows. If  $v$  is the vertex returned by such an oracle, then for every ancestor  $u$  of  $v$ , such that  $\kappa_u = \vee$ , and every  $w \in \text{Children}(v)$  that is not an ancestor of  $v$ , a number of recursive calls with  $\Phi_w$  and distance parameter significantly larger than  $\epsilon$  are used. The following Lemma implies that if for each of these vertices one of the recursive calls returned 0, then we know that  $\sigma \notin \text{SAT}(\Phi)$ .

**Lemma 5.3** The set of special relatives of  $v \in V$  is the set  $T$  of every  $u$  that is not an ancestor of  $v$  or  $v$  itself but is a child of an ancestor  $w$  of  $v$ , where  $\kappa_w \equiv \vee$ . If  $\sigma \notin \text{SAT}(\Phi_u)$  for every  $u \in T \cup \{v\}$ , then  $\sigma \notin \text{SAT}(\Phi)$ .

**Proof.** If  $\text{depth}_\Phi(v) = 0$  then  $\sigma \notin \text{SAT}(\Phi_v)$  implies  $\sigma \notin \text{SAT}(\Phi)$ . Assume by induction that the lemma holds for any formula  $\Phi' = (V', E', r', X', \kappa')$ , assignment  $\sigma'$  to  $\Phi'$  and vertex  $u \in V'$  such that  $0 \leq \text{depth}_{\Phi'}(u) < \text{depth}_\Phi(v)$ . Let  $w$  be the parent of  $v$ . Observe that the special relatives of  $w$  are a subset of the special relatives of  $v$  and hence by the induction assumption we only need to prove that  $\sigma \notin \text{SAT}(\Phi_w)$  in order to infer that  $\sigma \notin \text{SAT}(\Phi)$ .

If  $\kappa_w \equiv \wedge$ , then  $\sigma \notin \text{SAT}(\Phi_v)$  implies that  $\sigma \notin \text{SAT}(\Phi_w)$ . If  $\kappa_w \equiv \vee$ , then  $\sigma \notin \text{SAT}(\Phi_v)$  and  $\sigma \notin \text{SAT}(\Phi_u)$  for every  $u \in T$  implies that  $\sigma \notin \text{SAT}(\Phi_w)$ , since  $\text{Children}(w) \setminus \{v\} \subseteq T$ . ■

The following Lemma state that if  $\sigma$  is  $\epsilon$ -far from  $\text{SAT}(\Phi)$ , then  $(\epsilon, \sigma)$ -critical vertices are abundant, and so we can locate one of them by merely sampling a sufficiently large (linear in  $1/\epsilon$ ) number of vertices.

The main part of the proof that it holds is in showing that if  $\sigma$  is only  $2\epsilon/3$ -far from  $\text{SAT}(\Phi)$ , then there exists an  $(\epsilon, \sigma)$ -critical vertex for  $\sigma$ . We first show that this is sufficient to show the claimed abundance of  $(\epsilon, \sigma)$ -critical vertices, and then state and prove the required lemma.

**Lemma 5.4** If  $\sigma$  is  $\epsilon$ -far from  $\text{SAT}(\Phi)$ , then  $|\{v | v \text{ is } (\epsilon, \sigma)\text{-critical}\}| \geq \epsilon|\Phi|/4$ .

**Proof.** Set  $\text{Critical}_{\epsilon, \sigma} = \{v | v \text{ is } (\epsilon, \sigma)\text{-critical}\}$  and assume on the contrary that  $|\text{Critical}_{\epsilon, \sigma}| < \epsilon|\Phi|/4$ . Set  $\sigma'$  to be an assignment to  $X$  so that for every  $s \in V$  where  $\kappa_s \in X$ , we have that  $\sigma'(\kappa_s) = 1$  if  $\kappa_s \in \text{Critical}_{\epsilon, \sigma}$  and otherwise  $\sigma'(x) = \sigma(x)$ . Thus  $\text{Critical}_{\epsilon, \sigma'} = \emptyset$ . By the triangle inequality we have that  $\text{farness}(\sigma, \text{SAT}(\Phi)) - \text{farness}(\sigma', \text{SAT}(\Phi)) \leq \text{farness}(\sigma', \sigma)$ . Finally since  $\text{Critical}_{\epsilon, \sigma'} = \emptyset$ , Lemma 5.5, which we prove below, asserts that  $\text{farness}(\sigma', \text{SAT}(\Phi)) < 2\epsilon/3$  and we reach a contradiction. ■

**Lemma 5.5** *If there is no  $(\epsilon, \sigma)$ -critical vertex, then  $\sigma$  is  $2\epsilon/3$ -close to  $SAT(\Phi)$ .*

**Proof.** We shall show that if  $\sigma$  is  $2\epsilon/3$ -far from  $SAT(\Phi)$ , then there exists an  $(\epsilon, \sigma)$ -critical vertex. Assume that  $\sigma$  is  $2\epsilon/3$ -far from  $SAT(\Phi)$ . This implies that  $r$  is an  $(\epsilon, \sigma)$ -important vertex. Hence an  $(\epsilon, \sigma)$ -important vertex exists. Let  $v$  be an  $(\epsilon, \sigma)$ -important vertex such that  $\text{depth}_\Phi(v)$  is maximal. Consequently, none of the vertices in  $\text{Children}(v)$  is  $(\epsilon, \sigma)$ -important. We next prove that  $v$  is  $(\epsilon, \sigma)$ -critical.

Assume on the contrary that  $v$  is not  $(\epsilon, \sigma)$ -critical. Consequently  $\kappa_v \notin X$  and hence to get a contradiction it is sufficient to show that there exists an  $(\epsilon, \sigma)$ -important vertex in  $\text{Children}(v)$ . If  $\kappa_v \equiv \vee$ , then by Observation 2.17 we get that  $\text{farness}(\sigma, SAT(\Phi_{h(v)})) \geq (2\epsilon/3)(1 + 2\epsilon/3)^{\lfloor \text{depth}_\Phi(h(v))/3 \rfloor}$ , and hence  $h(v)$  is  $(\epsilon, \sigma)$ -important.

Assume that  $\kappa_v \equiv \wedge$ . Let  $u$  be such that  $\text{farness}(\sigma, SAT(\Phi_u)) \geq \text{farness}(\sigma, SAT(\Phi_v))$ . Observation 2.16 asserts that such a vertex exists. We assume that  $\text{depth}_\Phi(u) > 2$ , since otherwise it cannot be the case that  $\text{farness}(\sigma, SAT(\Phi_u)) < (2\epsilon/3)(1 + 2\epsilon/3)^0$ . Let  $w \in V$  be the parent of  $v$ . Since  $w$  is an ancestor of  $v$  it is  $(\epsilon, \sigma)$ -important, and hence  $\text{farness}(\sigma, SAT(\Phi_w)) \geq (2\epsilon/3)(1 + 2\epsilon/3)^{\lfloor \text{depth}_\Phi(w)/3 \rfloor}$ . Since  $\Phi$  is basic we have that  $\kappa_w \equiv \vee$ . Thus by Observation 2.17 we get that  $\text{farness}(\sigma, SAT(\Phi_v)) \geq (2\epsilon/3)(1 + 2\epsilon/3)^{1 + \lfloor \text{depth}_\Phi(w)/3 \rfloor}$ . Finally since  $\text{farness}(\sigma, SAT(\Phi_u)) \geq \text{farness}(\sigma, SAT(\Phi_v))$  and  $\text{depth}_\Phi(u) = \text{depth}_\Phi(w) + 2$  we get that  $\text{farness}(\sigma, SAT(\Phi_u)) \geq (2\epsilon/3)(1 + 2\epsilon/3)^{\lfloor \text{depth}_\Phi(u)/3 \rfloor}$ . ■

## 5.2 Algorithm

This algorithm detects far inputs with probability  $\Theta(\epsilon)$ , but this can be amplified to  $2/3$  using iterated applications.

---

### Algorithm 3

---

**Input:** read-once basic formula  $\Phi = (V, E, r, X, \kappa)$ , a parameter  $\epsilon > 0$ , oracle to  $\sigma$ .

**Output:**  $z \in \{0, 1\}$ .

```

1: if  $\epsilon > 1$  then return 1
2: if  $\kappa_r \in X$  then return  $\sigma(\kappa_r)$ 
3: Pick  $s$  uniformly at random from all  $v$  such that  $\kappa_v \in X$ 
4:  $A \leftarrow$  all ancestors  $v$  of  $s$  such that  $\kappa_v \equiv \vee$ 
5:  $R \leftarrow (\bigcup_{v \in A} \text{Children}(v)) \setminus \{w \mid w \text{ is an ancestor of } s\}$ 
6: if  $|R| > 3\epsilon^{-2} \log(2\epsilon^{-1})$  then return 1
7: for all  $u \in R$  do
8:    $y_u \leftarrow 1$ 
9:   for  $i = 1$  to  $\lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil$  do  $y_u \leftarrow y_u \wedge \text{Algorithm 3}(\Phi_u, \sigma, 4\epsilon/3)$ 
10: end for
11: return  $\sigma(\kappa_s) \vee \bigvee_{u \in R} y_u$ 

```

---

We can now proceed to prove the correctness of Algorithm 3. Algorithm 3 is clearly non-adaptive. We next prove that it always returns “1” for an assignment that satisfies the formula,



and returns “0” with probability linear in  $\epsilon$  for an assignment that is  $\epsilon$ -far from satisfying the formula. Using  $O(1/\epsilon)$  independent iterations amplifies the later probability to  $2/3$ .

**Lemma 5.6** *For  $\epsilon > 0$ , Algorithm 3 halts after using at most  $\epsilon^{-16+16 \log \epsilon}$  queries, when called with  $\Phi$ ,  $\epsilon$  and oracle access to  $\sigma$ .*

**Proof.** The proof is formulated as an inductive argument over the value of the (real) fairness parameter  $\epsilon$ . However, it is formulated in a way that it can be viewed as an inductive argument over the integer valued  $\lceil \log(\alpha \epsilon^{-1}) \rceil$ , for an appropriate global constant  $\alpha$ .

If  $\epsilon > 1$ , then the condition in Line 1 is satisfied, and there are no queries or recursive calls. Hence we assume that  $\epsilon \leq 1$ . Observe that in a specific instantiation at most one query is used, since a query is only made on Line 2 or on Line 11, and always as part of a “return” command. Hence the number of queries is upper bounded by the number of calls to Algorithm 3 (initial and recursive). We shall show that the number of these calls is at most  $\epsilon^{-16+16 \log \epsilon}$ .

Assume by induction that for some  $\eta \leq 1$ , for every  $\eta \leq \eta' \leq 1$ , every formula  $\Phi'$  and assignment  $\sigma'$  to  $\Phi'$ , on call to Algorithm 3 with  $\Phi'$ ,  $\eta'$  and an oracle to  $\sigma'$ , at most  $\eta'^{-16+16 \log \eta'}$  calls to Algorithm 3 are made (including recursive ones).

Assume that  $\epsilon > 3\eta/4$ . If  $\kappa_r \in X$ , then the condition on Line 1 is satisfied and hence there are no recursive calls. Thus Algorithm 3 is called only once and  $1 \leq \epsilon^{-16+16 \log \epsilon}$ .

Assume that  $\kappa_r \notin X$ . Note that every recursive call is done by Line 9. By Line 7 and Line 9 at most  $|R| \cdot \lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil$  recursive calls are done. The condition on Line 6 ensures that  $|R| \cdot \lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil \leq 3\epsilon^{-2} \log(2\epsilon^{-1}) \cdot \lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil$ . According to Line 9 each one of these recursive calls is done with distance parameter  $4\epsilon/3 > \eta$ . Thus by the induction assumption the number of calls to Algorithm 3 is at most  $1 + 3\epsilon^{-2} \log(2\epsilon^{-1}) \cdot \lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil \cdot (4\epsilon/3)^{-16+16 \log(4\epsilon/3)}$ . This is less than  $\epsilon^{-16+16 \log \epsilon}$ . ■

The following theorem will be immediate from Lemma 5.6 above when coupled with Lemma 5.9 and Lemma 5.11 below.

**Theorem 5.7** *Let  $\epsilon > 0$ . When Algorithm 3 is called with  $\Phi$ ,  $\epsilon$  and an oracle to  $\sigma$ , it uses at most  $\epsilon^{-16+16 \log \epsilon}$  queries; if  $\sigma \in \text{SAT}(\Phi)$  then it always returns 1, and if  $\sigma$  is  $\epsilon$ -far from  $\text{SAT}(\Phi)$  then it returns 0 with probability at least  $\epsilon/8$ .*

Theorem 5.7 does not imply that Algorithm 3 is an  $\epsilon$ -test for  $\text{SAT}(\Phi)$ . However it does imply that in order to get an  $\epsilon$ -test for  $\text{SAT}(\Phi)$  it is sufficient to do the following. Call Algorithm 3 repeatedly  $\lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil$  times, return 0 if any of the calls returned 0, and otherwise return 1. This only increases the query complexity to the value in the following corollary.

**Corollary 5.8** *There exists an  $\epsilon$ -test for  $\Phi$ , that uses at most  $\epsilon^{-20+16 \log \epsilon}$  queries.*

**Lemma 5.9** *Let  $\epsilon > 0$  and  $\sigma \in \text{SAT}(\Phi)$ . Algorithm 3 returns 1 when called with  $\Phi$ ,  $\epsilon$  and an oracle to  $\sigma$ .*

**Proof.** To prove the lemma we shall show that if Algorithm 3 returns 0, when called with  $\Phi$ ,  $\epsilon$  and oracle access to  $\sigma$ , then  $\sigma \notin \text{SAT}(\Phi)$ . If  $\text{depth}(\Phi) = 0$  then the condition in Line 1 is satisfied and  $\sigma(\kappa_r)$  is returned. Hence  $\sigma(\kappa_r) = 0$  and therefore  $\sigma \notin \text{SAT}(\Phi)$ . Assume that for every  $\epsilon' > 0$ ,

$\Phi'$  where  $\text{depth}(\Phi') < \text{depth}(\Phi)$ , and assignment  $\sigma'$  to  $\Phi'$ , if Algorithm 3 returns 0, when called with  $\Phi'$ ,  $\epsilon'$  and oracle access to  $\sigma'$ , then  $\sigma' \notin \text{SAT}(\Phi)$ .

Observe that the only other way a 0 can be returned is through Line 11, if it is reached. Let  $R$  be the set of vertices on which there was a recursive call in Line 9 and  $\kappa_s$  the variable whose value is queried on Line 11. According to Line 11 a 0 is returned if and only if  $\sigma(\kappa_s) = 0$  and for every  $u \in R$  there was at least one recursive call with  $\Phi_u$  and distance parameter  $4\epsilon/3$  that returned a 0. By the induction assumption this implies that  $\sigma \notin \text{SAT}(\Phi_u)$  for every  $u \in R$ . Note that the set  $R$  satisfies the exact same conditions that the set  $T$  satisfies in Lemma 5.3. Hence, Lemma 5.3 asserts that  $\sigma \notin \text{SAT}(\Phi)$ . ■

We now turn to proving soundness. This depends on first noting that the algorithm will indeed check the paths leading to critical vertices.

**Observation 5.10** *If the vertex  $s$  picked in Line 3 is  $(\epsilon, \sigma)$ -critical, then it will not trigger the condition of Line 6.*

**Proof.** Definition 5.1 in particular implies that for every  $u \in A$  (as per Line 4) we have  $|\text{Children}(u)| \leq (3/2\epsilon)(1 + 2\epsilon/3)^{-\lfloor \text{depth}_\Phi(u)/3 \rfloor} \leq 3/2\epsilon$ , as otherwise  $\sigma$  will be too close to satisfying  $\Phi_u$ . Also, from Observation 5.2 we know that  $\text{depth}_\Phi(s) \leq 4\epsilon^{-1} \log(2\epsilon^{-1})$  and so  $|A| \leq 2\epsilon^{-1} \log(2\epsilon^{-1}) + 1$ .

The two together give us the bound  $|R| \leq (3/2\epsilon - 1)(2\epsilon^{-1} \log(2\epsilon^{-1}) + 1) \leq 3\epsilon^{-2} \log(2\epsilon^{-1})$ , and so the condition in Line 3 is not triggered. ■

**Lemma 5.11** *Let  $\sigma$  be  $\epsilon$ -far from  $\text{SAT}(\Phi)$ . If Algorithm 3 is called with  $\epsilon$ ,  $\Phi$  and an oracle to  $\sigma$ , then it returns 0 with probability at least  $\epsilon/8$ .*

**Proof.** The base case,  $\kappa_r \in X$ , is handled correctly by Line 1. Assume next that  $\epsilon > 3/4$ . Assume that the vertex  $s$  selected in Line 3 is  $(\epsilon, \sigma)$ -critical. Hence by definition  $\sigma$  is more than  $1/2$ -far from  $\text{SAT}(\Phi_u)$  for every ancestor  $u$  of  $s$ . Thus by Observation 2.17 we have that  $\kappa_u \equiv \wedge$  for every ancestor  $u$  of  $s$ . Consequently, by Line 2 and Line 11 the value returned will be  $\sigma(\kappa_s)$ , and  $\sigma(\kappa_s) = 0$  because  $s$  is  $(\epsilon, \sigma)$ -critical. By Lemma 5.4, with probability at least  $3/16$  the vertex  $s$  selected in Line 3 is  $(\epsilon, \sigma)$ -critical. Thus, 0 is returned with probability at least  $3/16$ , which is greater than  $\epsilon/8$  when  $3/4 < \epsilon \leq 1$ .

For all other  $\epsilon$  we proceed by induction over the depth. Assume that for any formula  $\Phi'$  such that  $\text{depth}(\Phi') < \text{depth}(\Phi)$  and any assignment  $\sigma'$  to  $\Phi'$  that is  $\eta$ -far from  $\text{SAT}(\Phi')$  (for any  $\eta$ ), Algorithm 3 returns 0 with probability at least  $\eta/8$ . Given this we prove that 0 is returned with probability at least  $\epsilon/8$  for  $\Phi$  and  $\sigma$ .

Assume first that the vertex  $s$  selected in Line 3 is  $(\epsilon, \sigma)$ -critical. Let  $A, R$  be the sets from Line 5 and Line 4. Since  $s$  is  $(\epsilon, \sigma)$ -critical, by definition for every  $u \in A$  we have that  $\sigma$  is  $2\epsilon/3$ -far from  $\text{SAT}(\Phi_u)$ . Also, because  $s$  is  $(\epsilon, \sigma)$ -critical, by definition for every  $u \in A$  and  $w \in \text{Children}(u) \cap R$  we have that  $w \neq h(u)$ , and therefore by Observation 2.17 we have that  $\sigma$  is  $4\epsilon/3$ -far from  $\text{SAT}(\Phi_w)$  for every  $w \in R$ . By the induction assumption, for every  $w \in R$ , with probability at least  $1 - (4\epsilon/3)/8$  Algorithm 3 returns 0 when called with  $4\epsilon/3$ ,  $\Phi_w$  and an oracle to  $\sigma$ . Hence, for every  $w \in R$ , the probability that on  $\lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil$  such independent calls to Algorithm 3 the value 0 was never returned is at most  $(1 - (4\epsilon/3)/8)^{\lceil 20\epsilon^{-1} \log \epsilon^{-1} \rceil}$ . This is less than  $(\epsilon^{-2} \log(2\epsilon^{-1}))/6$ . Observation

5.10 ensures that  $|R| \leq 3\epsilon^{-2} \log(2\epsilon^{-1})$ , and in particular the condition in Line 6 is not invoked and the calls in Line 9 indeed take place. By the union bound over the vertices of  $R$ , with probability at least  $1/2$ , for every  $u \in R$  at least one of calls to Algorithm 3 with  $4\epsilon/3$ ,  $\Phi_u$  and an oracle to  $\sigma$  returned the value 0. This means that for every  $u \in R$ ,  $y_u$  in Line 9 was set to 0 and remained 0. Consequently this is the value returned in Line 11

Finally, since  $\sigma$  is  $\epsilon$ -far from  $SAT(\Phi)$ , by Lemma 5.4 the vertex  $s$  selected in Line 3 is  $(\epsilon, \sigma)$ -critical with probability at least  $\epsilon/4$ . Therefore 0 is returned with probability at least  $\epsilon/8$ . ■

## 6 The Computational Complexity of the Testers and Estimator

There are two parts to analyzing the computational complexity of a test for a massively parametrized property. The first part is the running time of the preprocessing phase, which reads the entire parameter part of the input, in our case the formula, but has no access yet to the tested part, in our case the assignment. This part is subject to traditional running time and working space definitions, and ideally should have a running time that is quasi-linear or at least polynomial in the parameter size.

In our case, the preprocessing part would need to take a  $k$ -ary formula and convert it to the basic form corresponding to the algorithm that we run. We would also need to put the formula in a data structure that allows the following operations to be performed as quickly as possible:

For Algorithm 1, we would need to quickly pick a child of a vertex with probability proportional to its sub-formula size, and know who are the light children as well as what is the relative size of the smallest child. This mainly requires storing the size of every sub-formula for every vertex of the tree, as well as sorting the children of each vertex by their sizes and storing the value of the corresponding “ $\ell$ ”. Algorithm 2 requires very much the same additional data as Algorithm 1.

For Algorithm 3, we would need to navigate the tree both downwards and upwards (for finding the ancestors of a vertex), as well as the ability to pick a vertex corresponding to a variable at random, which in itself does not require special preprocessing.

It is not hard to see that both constructing the normal form and calculating the above additional data can be done very efficiently. Furthermore, the only part that depends on epsilon is designating the light children, but this can also be done “for all epsilon” at a low cost (by storing the range of  $\epsilon$  for every positive  $\ell$ ).

The second part is analyzing the running time complexity of the algorithm. Once the above preprocessing is performed, the time per instantiation (and thus per query) of the algorithm will be very small (where we charge the time it takes to calculate a recursive call to the recursive instantiation). This would make it a cost logarithmic in the input size per query (multiplied by the time it takes to write and read an address) – where the log incurrence is in fact only when we need to randomly choose a child according to its weight.

## 7 An Untestable 4-Valued Formula

We describe here a read-once formula over an alphabet with 4 values, defining a property that cannot be  $1/4$ -tested using a constant number of queries. The formula will have a very simple

structure, with only one gate type. However this gate satisfies no monotonicity-like property with respect to the alphabet.

## 7.1 The formula

For convenience we denote our alphabet by  $\Sigma = \{0, 1, P, F\}$ . An input is said to be *accepted* by the formula if, after performing the calculations in the gates, the value received at the root of the tree is not “F”. We restrict the input variables to  $\{0, 1\}$ , although it is easy to see that the following argument holds also if we allow other values to the input variables (and also if we change the acceptance condition to the value at the root having to be “P”).

**Definition 7.1** *The balancing gate is the gate that receives two inputs from  $\Sigma$  and outputs the following.*

- For  $(0, 0)$  the output is 0 and for  $(1, 1)$  the output is 1.
- For  $(1, 0)$  and  $(0, 1)$  the output is P.
- For  $(P, P)$  the output is P,
- For anything else the output is F.

For a fixed  $h > 0$ , the balancing formula of height  $h$  is the formula defined by the following.

- The tree is the full balanced binary tree of height  $h$ , and hence there are  $2^h$  variables.
- All gates are set to the balancing gate.
- The formula accepts if the value output at the root is not “F”.

We denote the variables of the formula in their order by  $x_0, \dots, x_{2^h-1}$ . The following is easy.

**Lemma 7.2** *An assignment  $a_0 \in \{0, 1\}, \dots, a_{2^h-1} \in \{0, 1\}$  to  $x_0, \dots, x_{2^h-1}$  is accepted by the formula if and only if for every  $0 < k \leq h$  and every  $0 \leq i < 2^{h-k}$ , the number of 1 values in  $a_{i2^k}, \dots, a_{(i+1)2^k-1}$  is either 0,  $2^k$  or  $2^{k-1}$ .*

**Proof.** Denote the number of 1 values in variables descending from a gate  $u$  by  $\sharp_1(u)$ . Let us prove by induction on  $k$  that  $\sharp_1(v) = 0$  if and only if the value of  $v$  is 0,  $\sharp_1(v) = 2^k$  if and only if the value of  $v$  is 1, and  $\sharp_1(v) = 2^{k-1}$  if and only if the value of  $v$  is P. For  $k = 1$  we have the two inputs of  $v$ , and by the definition of the balancing gate the claim follows.

For  $k > 1$ , we have  $2^k$  variables which are all descendants of the same gate  $v$ . By the induction hypothesis, for both children of  $v$ , denoted  $u, w$ , we have that  $\sharp_1(u), \sharp_1(w) \in \{0, 2^{k-2}, 2^{k-1}\}$  and that this determines their value. If  $\sharp_1(w) = \sharp_1(u) = 0$  then they both evaluate to 0 and so does  $v$ . Similarly  $\sharp_1(w) = \sharp_1(u) = 2^{k-1}$  then both evaluate to 1 and so does  $v$ . If  $\sharp_1(u) = 2^{k-1}$  and  $\sharp_1(w) = 0$ , then  $u$  evaluates to 1 and  $w$  to 0 and indeed  $v$  evaluates to P (similarly for the symmetric case). The remaining case is  $\sharp_1(u) \in \{0, 2^{k-1}\}$  and  $\sharp_1(w) = 2^{k-2}$  (and the symmetric case), by the induction hypothesis and the definition of the balancing gate this implies that  $v$  evaluates to F and the formula is unsatisfied. ■

In other words, every “binary search interval” is either all 0, or all 1, or has the same number of 0 and 1. This will allow us to easily prove that certain inputs are far from satisfying the property.

## 7.2 Two distributions

We now define two distributions, one over satisfying inputs and the other over far inputs.

**Definition 7.3** *The distribution  $D_Y$  is defined by the following process.*

- Uniformly pick  $2 \leq k \leq h$ .
- For every  $0 \leq i < 2^{h-k}$ , independently pick either  $(y_{i,0}, y_{i,1}) = (0, 1)$  or  $(y_{i,0}, y_{i,1}) = (1, 0)$  (each with probability  $1/2$ ).
- For every  $0 \leq i < 2^{h-k}$ , set

$$x_{i2^k} = \cdots = x_{i2^k+2^{k-1}-1} = y_{i,0}; \quad x_{i2^k+2^{k-1}} = \cdots = x_{(i+1)2^k-1} = y_{i,1}.$$

**Definition 7.4** *The distribution  $D_N$  is defined by the following process.*

- Uniformly pick  $2 \leq k \leq h$ .
- For every  $0 \leq i < 2^{h-k}$ , independently choose  $(z_{i,0}, z_{i,1}, z_{i,2}, z_{i,3})$  to have either one 1 and three 0 or one 0 and three 1 (each of the 8 possibilities with probability  $1/8$ ).
- For every  $0 \leq i < 2^{h-k}$ , set

$$\begin{aligned} x_{i2^k} = \cdots = x_{i2^k+2^{k-2}-1} &= z_{i,0}; & x_{i2^k+2^{k-2}} = \cdots = x_{i2^k+2^{k-1}-1} &= z_{i,1}; \\ x_{i2^k+2^{k-1}} = \cdots = x_{i2^k+2^{k-1}+2^{k-2}-1} &= z_{i,2}; & x_{i2^k+2^{k-1}+2^{k-2}} = \cdots = x_{(i+1)2^k-1} &= z_{i,3}. \end{aligned}$$

It is easier to illustrate this by considering the calculation that results from the distributions. In both distributions we can think of a randomly selected level  $k$  (counted from the bottom, where the leaf level 0 and the level above it are never selected). In  $D_Y$ , the output of all gates at or above level  $k$  is “P”, while the inputs to every gate at level  $k$  will be either  $(0, 1)$  or  $(1, 0)$ , chosen uniformly at random.

In  $D_N$  all gates at level  $k$  will output “F” (note however that we cannot query a gate output directly); looking two levels below, every gate as above holds the result of a quadruple chosen uniformly from the 8 choices described in the definition of  $D_N$  (the quadruple  $(z_{i,0}, z_{i,1}, z_{i,2}, z_{i,3})$ ). At level  $k-2$  or lower the gate outputs are 0 and 1 and their distribution resembles very much the distribution as in the case for  $D_Y$  – as long as we can not “focus” at the transition level  $k$ . This is formalized in terms of lowest common ancestors below.

**Lemma 7.5** *Let  $Q \subset \{1, \dots, 2^h\}$  be a set of queries, and let  $H \subset \{0, \dots, h\}$  be the set of levels containing lowest common ancestors of subsets of  $Q$ . Conditioned on neither  $k$  nor  $k-1$  being in  $H$ , both  $D_Y$  and  $D_N$  induce exactly the same distribution over the outcome of querying  $Q$ .*

**Proof.** Let us condition the two distributions on a specific value of  $k$  satisfying the above. For two queries  $q, q' \in Q$  whose lowest common ancestor is on a level below  $k-1$ , with probability 1 they will receive the exact same value (this holds for both  $D_N$  and  $D_Y$ ). The reason is clear from the construction – their values will come from the same  $y_{i,j}$  or  $z_{i,j}$ .

Now let  $Q'$  contain one representative from every set of queries in  $Q$  that must receive the same value by the above argument. For any  $q, q' \in Q'$ , their lowest common ancestor is on a level above  $k$ . For  $D_Y$  it means that  $x_q$  takes its value from some  $y_{i,j}$  and  $x_{q'}$  takes its value from some  $y_{i',j'}$  where  $i \neq i'$ . Because each pair  $(y_{i,0}, y_{i,1})$  is chosen independently from all others, this means that the outcome of the queries in  $Q'$  is uniformly distributed among the  $2^{|Q'|}$  possibilities. The same argument (with  $z_{i,j}$  and  $z_{i',j'}$  instead of  $y_{i,j}$  and  $y_{i',j'}$ ) holds for  $D_N$ . Hence the distribution of outcomes over  $Q'$  is the same for both distributions, and by extension this holds over  $Q$ . ■

On the other hand, the two distributions are much different with respect to satisfying the formula.

**Lemma 7.6** *An input chosen according to  $D_Y$  always satisfies the balancing formula, while an input chosen according to  $D_N$  is always  $1/4$ -far from satisfying it.*

**Proof.** This follows immediately from the construction and Lemma 7.2. ■

### 7.3 Proving non-testability

We use here the following common application of Yao's method (see e.g. [6]).

**Lemma 7.7** *If  $D_Y$  is a distribution over satisfying inputs and  $D_N$  a distribution over  $\epsilon$ -far ones, such that for any fixed set of queries  $Q$  with  $|Q| \leq l$  the probability distributions over the outcomes differ by less than  $\frac{1}{3}$  (in the variation distance norm) for  $D_Y$  and  $D_N$ , then there is no non-adaptive  $\epsilon$ -test for the property that makes at most  $l$  queries (1-sided or 2-sided).*

This allows us to conclude the proof.

**Theorem 7.8** *Testing for being a satisfying assignment of the balancing formula of height  $h$  requires at least  $\Omega(h)$  queries for a non-adaptive test and  $\Omega(\log h)$  queries for a possibly adaptive one.*

**Proof.** We note that for any set of queries  $Q$ , the size of the set of lowest common ancestors (outside  $Q$  itself) is less than  $Q$ , and hence (in the notation of Lemma 7.5) we have  $|H| \leq |Q|$ . Now if  $|Q| = o(h)$ , then the event of Lemma 7.5 happens with probability  $1 - o(1)$ , and hence the variation distance between the two (unconditional) distributions over outcomes is  $o(1)$ . Together with Lemma 7.6 this fulfills the conditions for Lemma 7.7 for concluding the proof.

For adaptive algorithms the bound follows by the standard procedure that makes an adaptive algorithm into a non-adaptive one at an exponential cost (by querying in advance the algorithm's entire decision tree given its internal coin tosses). ■

## References

- [1] Noga Alon, Michael Krivelevich, Ilan Newman, and Mario Szegedy. Regular languages are testable with a constant number of queries. *SIAM J. Comput.*, 30(6):1842–1862, 2000.
- [2] Eli Ben-Sasson, Prahladh Harsha, Oded Lachish, and Arie Matsliah. Sound 3-query pcpps are long. *ACM Trans. Comput. Theory*, 1:7:1–7:49, September 2009.

- [3] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. Some 3CNF properties are hard to test. *SIAM J. Comput.*, 35(1):1–21, 2005.
- [4] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.
- [5] Sourav Chakraborty, Eldar Fischer, Oded Lachish, Arie Matsliah, and Ilan Newman. Testing  $st$ -connectivity. In *APPROX-RANDOM*, pages 380–394, 2007.
- [6] Eldar Fischer. The art of uninformed decisions: A primer to property testing. *Current Trends in Theoretical Computer Science: The Challenge of the New Century*, I:229–264, 2004.
- [7] Eldar Fischer, Oded Lachish, Ilan Newman, Arie Matsliah, and Orly Yahalom. On the query complexity of testing orientations for being eulerian. *Transactions on Algorithms*, to appear.
- [8] Eldar Fischer, Ilan Newman, and Jiri Sgall. Functions that have read-twice constant width branching programs are not necessarily testable. *Random Struct. Algorithms*, 24(2):175–193, 2004.
- [9] Eldar Fischer and Orly Yahalom. Testing convexity properties of tree colorings. *Algorithmica*, 60(4):766–805, 2011.
- [10] Oded Goldreich. A brief introduction to property testing. In Oded Goldreich, editor, *Property Testing*, pages 1–5. Springer-Verlag, 2010.
- [11] Oded Goldreich, Shaffi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45:653–750, July 1998.
- [12] Shirley Halevy, Oded Lachish, Ilan Newman, and Dekel Tsur. Testing orientation properties. *Electronic Colloquium on Computational Complexity (ECCC)*, (153), 2005.
- [13] Shirley Halevy, Oded Lachish, Ilan Newman, and Dekel Tsur. Testing properties of constraint-graphs. In *IEEE Conference on Computational Complexity*, pages 264–277, 2007.
- [14] Ilan Newman. Testing membership in languages that have small width branching programs. *SIAM J. Comput.*, 31(5):1557–1570, 2002.
- [15] Ilan Newman. Property testing of massively parametrized problems - a survey. In Oded Goldreich, editor, *Property Testing*, pages 142–157. Springer-Verlag, 2010.
- [16] Dana Ron. Property testing: A learning theory perspective. *Found. Trends Mach. Learn.*, 1:307–402, March 2008.
- [17] Dana Ron. *Algorithmic and Analysis Techniques in Property Testing*. 2010.
- [18] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.